0.1 Multilayer Feedforward Networks are Universal Approximators

In 1988 Maxwell Stinchcombe and Halber White show that Multilayer Feedforward Networks with one hidden layer are Universal Approximators.

The proof is based on that arbitrary squashing function are capable of approximating any Borel measurable function from one finite dimensional space to another to any desired degree of accuracy, provided sufficiently hidden units are available. In this sense Feedforward networks are a class of Universal approximators. (REWRITE THREE LAST SENTENCES)

Firstly we are going to start with some basic definitions.

Definition 0.1.1. Affine functions

Affine functions are defined as all the function of the form $A(x) = w \cdot x + b$ from $\mathbb{R}^d$ to $\mathbb{R}$ where $d \in \mathbb{N} - \{0\}$, w and x are vector in $\mathbb{R}^d$, $\cdot$ denotes the usual dot product of vector, and $b \in \mathbb{R}$ is a scalar.

Definition 0.1.2. Squashing function

A function $\psi : \mathbb{R} \longrightarrow [0, 1]$ is a squashing function if

- It is non-decreasing.
- $\lim_{x\to\infty} \psi(x) =$.
- $\lim_{x\to-\infty} \psi(x) = 0$.

Squashing functions are measurable due to the fact that have at most countably discontinuities.

Most common examples are:

- Threshold functions.[1]
- Indicator functions: $\psi(\lambda) = 1_{\{\lambda>0\}}$.
- The ramp function: $\psi(\lambda) = \lambda 1_{\{0\leq\lambda\leq1\}} + 1_{\{\lambda>1\}}$
- The cosine squasher of Gallant and White (1988)

$$\psi(lambda) = (1 + cos(\lambda + 3\frac{\pi}{2})\frac{1}{2})1_{\{\frac{-\pi}{2}\leq\lambda\leq\frac{\pi}{2}\}}1_{\{\frac{\pi}{2}<\lambda\}}$$

Definition 0.1.3. Now we are going to define the familiar class of function of autput functions for a single hidden layer feedforward.

For any Borel measurable function $\mathcal{G}(\cdot)$ mapping $\mathbb{R}$ to $\mathbb{R}$ and $r \in \mathbb{N} - \{0\}$, let define the class of function

[1] A threshold function is a function that takes the value 1 is a specified function of the arguments exceeds a given threshold and 0 otherwise. [?]

$$\sum^{r}(\mathcal{G}) = \{f : \mathbb{R}^r \longrightarrow \mathbb{R} =$$

$$\sum_{j=1}^{q}(\prod_{k=1}^{l_j} \mathcal{G}(A_{jk}(x)), x \in \mathbb{R}^r, \beta_j \in \mathbb{R}, A_{jk} \in A^r, l_j \in \mathbb{N} - \{0\})\}$$